

EEB 581, Problem Set Three Solutions

1 : Suppose you are studying the number of visitations of a pollinator to a flower. Your hypothesis is that yellow flowers are better than red flowers (in terms of pollinator attraction). Previous studies have found that the number of visitors to red flowers follows a normal distribution with a mean of 200 visits per flower and a variance of 50. Suppose in a sample of 20 yellow flowers that the mean number of visits is 202 with a known variance (of visits per flower) of 50. Again, assume the number of visitors is normally distributed.

- (a) What is the probability of this data under the null hypothesis (that yellow and red flowers are equivalent)?

Let $T = 202$ be the mean number of visits. The mean and variance (under the null hypothesis) are $\mu_0 = 200$, and $\sigma_0^2 = 50/20$ (so that $f_0^2 = 50$). Under the null

$$\frac{T - 200}{\sqrt{50/20}} \sim U, \quad \text{implying} \quad P\left(U > \frac{202 - 200}{\sqrt{50/20}}\right) = P(U > 1.265) = 0.103$$

- (b) What is the critical value for a (one-sided) test of the null hypothesis at the $\alpha = 0.05$ level?

$$T_c(0.05) = \mu_0 + z_{(1-0.05)}\sqrt{50/20} = 200 + 1.64\sqrt{50/20} = 202.6$$

- (c) What are the values for (a) and (b) when the variance for yellow flowers (50) is instead a SAMPLE variance (i.e., an estimate of the true variance)? Hint: Would you now use a normal or a t distribution?

$$\frac{T - 200}{\sqrt{50/20}} \sim t_{19}, \quad \text{implying} \quad P\left(t_{19} > \frac{202 - 200}{\sqrt{50/20}}\right) = P(t_{19} > 1.265) = 0.1105991$$

in R, use `T<-(202-200)/sqrt(50/20) 0.5; 1-pt(T,19)`.

To find $t_{19,0.975}$ where $P(t_{19} > t_{19,0.95}) = 0.5$, use (in R) `qt(0.95,19)` which returns 1.73.

Hence, the critical value becomes

$$T_c(0.05, 19) = \mu_0 + t_{19,0.95}\sqrt{50/20} = 202.73$$

- (d) Suppose that yellow flowers are indeed better. Given the sample size (20) and assuming the variance (50) is the true value, how small an effect can we detect using a (one-sided) test of significance of $\alpha = 0.05$ with 80% power?

The power is

$$\Pr(T > 202.6) =$$

$$\Pr\left(\frac{t - \mu_1}{\sigma_1} > \frac{202.6 - \mu_1}{\sigma_1}\right) = \Pr\left(U > \frac{202.61 - \mu_1}{\sqrt{50/20}}\right) = 0.80$$

$\Pr(U > x) = 0.8$ implies $\Pr(U \leq x) = 0.2$ or that $x = -0.8416212$ (using `qnorm(0.2)`). Hence to get 80% power,

$$\frac{202.6 - \mu_1}{\sqrt{50/20}} = -0.841621, \quad \text{or} \quad 202.6 - \mu_1 = -0.841621 \cdot \sqrt{50/20}$$

or that $\mu_1 = 203.1 + 0.841621 \cdot \sqrt{50/20} = 203.9307$

- (e) Repeat the calculation in (d) assuming that the variance (50) is now an estimated value, not necessarily the true value.

We now use the t distribution,

$$\Pr\left(t_{19} > \frac{202.7 - \mu_1}{\sqrt{50/20}}\right) = 0.80$$

using **R**, note that `qt(0.2, 19)` returns **-0.861**, so that

$$\Pr(t_{19} > -0.861) = 1 - \Pr(t_{19} \leq -0.861) = 0.80$$

Hence, for 80% power,

$$\frac{202.7 - \mu_1}{\sqrt{50/20}} = -0.861, \quad \text{or} \quad \mu_1 = 202.7 + 0.861 \cdot \sqrt{50/20} = 204.06$$

- (f) Suppose the true mean and variance for yellow flowers are 201 and 10. How large a sample size is required to have a power of 80 percent of detecting a difference between red and yellow using a test of significance with level $\alpha = 0.05$? Compute this for both the normal (variance assumed known) and t (variance estimated) settings.

Variance assumed known (Normal distribution)

Applying Equation A5.4b and recalling for 80 percent power, we use $z_{(1-\beta)} = z_{(0.8)} = 0.842$ (as β is the probability of a type II error, so that $1 - \beta$ is the power) and likewise for $\alpha = 0.05$, $z_{(1-\alpha)} = 1.645$,

$$n = \left(\frac{z_{(1-\beta)} f_1 + z_{(1-\alpha)} f_0}{\mu_1 - \mu_0}\right)^2 = \left(\frac{0.842 \sqrt{10} + 1.645 \sqrt{50}}{1}\right)^2 = 204$$

Variance estimated (t distribution), here values for $z_{1-\alpha}$ are replaced with $t_{n-1, 1-\alpha}$

$$n = \left(\frac{t_{(n-1, 1-\beta)} f_1 + t_{(n-1, 1-\alpha)} f_0}{\mu_1 - \mu_0}\right)^2 = \left(\frac{0.861 \sqrt{10} + 1.729 \sqrt{50}}{1}\right)^2 = 223.5$$

- (g) If the true variance for yellow is 35, what is the probability that we observe a sample variance of 50 (or larger) given our sample size of 20.

Recalling that $\sum^n (x - \bar{x}) \sim \sigma^2 \chi_{n-1}^2$, we have

$$\text{Var} = \frac{1}{19} \sum (x - \bar{x}) \sim \frac{1}{19} \sigma^2 \chi_{19}^2 = 1.842 \chi_{19}^2$$

Hence

$$\Pr(\text{Var} > 50) = \Pr(1.842 \chi_{19}^2 > 50) = \Pr(\chi_{19}^2 > 27.14) = 0.10$$

In **R**, `1-pchisq(27.14, 19)`